
ANOVA Analysis of Variance



The Lean Six Sigma Institute

www.lean-six-sigma.pro

Reminders

- ❖ **Type I error:** a true null hypothesis incorrectly rejected
- ❖ **Variance:** The variance is one of several indices of variability that statisticians use to characterize the dispersion among the measures in a given population. To calculate the variance of a given population, it is necessary to first calculate the mean of the scores, then measure the amount that each score deviates from the mean and then square that deviation (by multiplying it by itself). Numerically, the variance equals the average of the several squared deviations from the mean
- ❖ **t-test:** a measure on a random sample (or pair of samples) in which a mean (or pair of means) appears in the numerator and an estimate of the numerator's standard deviation appears in the denominator based on the calculated s^2 or s^2 of the samples. If these calculations yield a value of (t) that is sufficiently different from zero, the test is considered to be statistically significant
- ❖ **Null hypothesis:** depending on the data, the null hypothesis either will or will not be rejected as a viable possibility. The null hypothesis is often the reverse of what the experimenter actually believes; it is put forward to allow the data to contradict it. If the test is significant, the null hypothesis is rejected.

Reminders

❖ The **Central Limit Theorem** consists of three statements:

- The mean of the sampling distribution of means is equal to the mean of the population from which the samples were drawn.
- The variance of the sampling distribution of means is equal to the variance of the population from which the samples were drawn divided by the size of the samples.
- If the original population is distributed normally (i.e. it is bell shaped), the sampling distribution of means will also be normal. If the original population is not normally distributed, the sampling distribution of means will increasingly approximate a normal distribution as sample size increases. (i.e. when increasingly large samples are drawn)

❖ The **standard error of the mean of a sampled distribution S_M** relates to the **population standard deviation σ** where n is the sample size, through the equations:

$$\sigma^2 = nS_M^2$$

$$S_M = \frac{\sigma}{\sqrt{n}}$$

Reminders

❖ **MSE** : the **Mean Square Error** is the average of **a** sample variances, **s**: from the Central Limit Theorem, the MSE tends to the population standard deviation

$$MSE = \frac{1}{a} \sum s^2 \approx \sigma^2$$

❖ **MSB** : the **Mean Square Between** is the variance of the distribution of the sample means multiplied by the sample size: from the Central Limit Theorem, the MSB also tends to the population standard deviation

$$MSB = ns_M^2 \approx \sigma^2$$

What is ANOVA for?

Allows testing of the difference between two or more means without increasing Type I error

Why don't we just use the t-test?

The t-test used to compare more than two different different means would lead to inflation of Type I error.

How ANOVA works

If the null hypothesis is true, then MSE and MSB should be about the same, because they both are estimates of the same population variance, σ^2 : in this case the ratio MSB/MSE is close to 1

If the null hypothesis is false, MSB will be larger than MSE, and the ratio MSB/MSE is greater than 1

The F statistic, compared with MSB/MSE, gives statistical significance to how much larger MSB is than MSE, permitting the null hypothesis to be evaluated probabilistically.

Example

Three groups of people sit an exam under three different room temperatures, hot-1, mild-2 and cold-3: ANOVA is used to verify if their scores are statistically different

- The null hypothesis is that the means of the scores are not different, and that:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

- This test has one **factor**, room temperature, and three **levels**, 1,2 and 3, denoted by the letter **a**:

$$a = 3$$

- The numbers of persons per group are designated as **n₁**, **n₂** and **n₃** (or **n_a**): we will assume here that the sample sizes in the **a** samples, are equal.

Example (cont.)

From the test results, we first estimate the population variance σ^2 by calculating the variances within the samples, to give the Mean Square Error, MSE.

- The population variance σ^2 is estimated by the average of the 3 (**a**) sample variances S^2 because, to apply the Central Limit Theorem, ANOVA assumes normally distributed populations, homogeneous variances and random, independent sampling.

$$MSE = \frac{1}{a} \sum s^2 \approx \sigma^2$$

$$MSE = \frac{1}{3} (58 + 69.8 + 47.6) = 58.5 \approx \sigma^2$$

	Hot	Mild	Cold
1	55	75	67
2	60	65	53
3	51	80	65
4	65	75	49
5	72	67	54
6	65	68	61
7	55	77	65
8	72	83	72
9	68	67	63
10	60	56	64
11	75	65	54
12	67	83	65
X bar	63,8	71,8	61,0
s ²	58,0	69,8	47,6

Example (cont.)

We then estimate the population variance σ^2 by calculating the variance of the sample means and multiplying by the sample size to give the Mean Square Between, MSB:-

$$MSB = ns_M^2 \approx \sigma^2$$

and where M is the grand mean of all the data:-

$$s_M^2 = \frac{\sum (\bar{X} - M)^2}{a-1}$$

$$s_M^2 = \frac{(63.8 - 65.5)^2 + (71.8 - 65.5)^2 + (61 - 65.5)^2}{3-1} = 31.2$$

$$MSB = ns_M^2 = 12 \times 31.2 = 374.4$$

	Hot	Mild	Cold	
1	55	75	67	
2	60	65	53	
3	51	80	65	
4	65	75	49	
5	72	67	54	
6	65	68	61	
7	55	77	65	
8	72	83	72	
9	68	67	63	
10	60	56	64	
11	75	65	54	
12	67	83	65	
X bar	63,8	71,8	61,0	M 65,5
s ²	58,0	69,8	47,6	

Example (cont.)

The ratio of MSB to MSE gives the F statistic, which, if large enough, allows the null hypothesis that the population means are equal, to be rejected.

If the null hypothesis is true, then both MSB and MSE estimate the same quantity, σ^2 , and the ratio becomes 1.

The sampling distribution of F shows the probability of calculating the F statistic given by the ratio, or larger. If the probability is lower than the significance level, the null hypothesis can be rejected.

Example (cont.)

The ratio is compared with the value found in the F table by using the degrees of freedom:-

$$F = MSB/MSE = 374.4/58.5 = 6.4$$

$$df_{\text{numerator}} = a-1 = 3-1 = 2$$

$$df_{\text{denominator}} = N-a = 36-3 = 33$$

For 2 degrees of freedom in the numerator, and 33 in the denominator, at a probability level of 0.05, the critical F value is 3.29 : at 6.4, we cannot reject the null hypothesis, so the means are statistically different

Example Minitab output (cont.)

The second column in the Minitab printout gives the intermediate SS (sum of the squares) calculation, which, like the degrees of freedom, adds up.

$$SS_{group} = n \sum (\bar{X} - M)^2 = 748.5$$

$$SS_{error} = \sum (X - \bar{X})^2 = 1930.5$$

$$SS_{total} = \sum (X - M)^2 = 2679$$

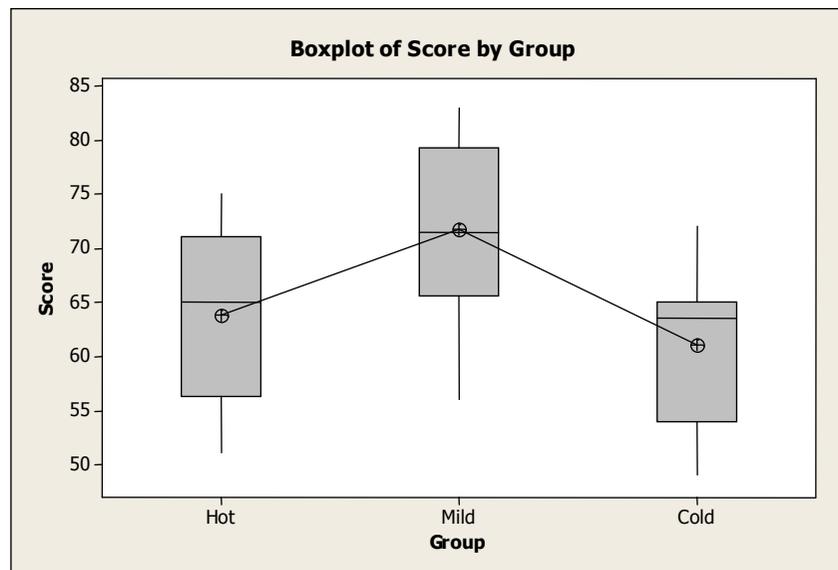
Example Minitab output (cont.)

- The sums of the squares add up: they divide the total sum of the squares into two components:
 - ❖ SSB, the group sum of the squares due to treatments
 - ❖ SSE, the error sum of the squares not due to treatments
- The MS (mean squares) column are estimates of variance given by SS/DF
- The F ratio comes from $MS \text{ group} / MS \text{ error}$
- The p value is the probability of obtaining an F value as large or larger than the one computed in the data assuming the null hypothesis is true.
- At the base, Minitab visually represents the data

Example Minitab output (cont.)

We conclude that at a probability of 0.004 we cannot assume that the null hypothesis is true, therefore we reject it and assume that the samples come from different populations.

We would state that: 'an analysis of variance was conducted showing that the temperature effect was significant: $F(2,33) = 6.4$, $p = 0.004$ ' and we might add a box plot of the data.



Example Minitab output (cont.)

Problem: we have no statistical measure of which population mean differs from the others, so Minitab offers significance tests which limit Type I error and are sufficiently powerful, such as:

- Fisher's LSD (least significant difference) which tests the null hypothesis that all the population means are equal;
- Tukey's HSD (honestly significantly different) testing all pairwise comparisons among means using the studentized t range;
- Dunnet's test which computes a t test between each experimental group and the control group.

General formulation of ANOVA

As the total sum of the squares adds the group and error sums of the squares, each score can be assumed to be the effect of three components:

- The overall population mean, which is the same as the overall mean of the group means, designated $\bar{\mu}$
- The treatment effect of being in the the group, designated α which is simply the difference between the group mean and the population mean $\bar{\mu} - \mu_j$
- The error, designated ε , which is the sum of all other effects, is normally distributed, and has a mean of zero

For the i^{th} subject in the j^{th} group:

$$y_{ij} = \bar{\mu} + \alpha_j + \varepsilon_{ij}$$

This Training Manual and all materials, procedures and systems herein contained or depicted (the "Manual") are the sole and exclusive property of La Rémige S.A.R.L.

The contents hereof contain proprietary information and materials that are the private property of La Rémige S.A.R.L. Unauthorized use, disclosure, or reproduction of any kind of any material contained in this Manual is expressly prohibited. The contents hereof are to be returned immediately upon termination of any relationship or agreement giving user authorization to possess or use such information or materials. Any unauthorized or illegal use shall subject the user to all remedies, both legal and equitable, available to La Rémige S.A.R.L. This Manual may be altered, amended or supplemented by La Rémige S.A.R.L. from time to time. In the event of any inconsistency or conflict between a provision in this Manual and any national, federal, provincial, state or local statute, regulation, order or other law, such law will supersede the conflicting or inconsistent provision(s) of this Manual in all properties subject to that law.

© 2008 by La Rémige S.A.R.L. All Rights Reserved.

"Lean Six Sigma" is a registered mark with INPI in France